Theory of Doping and Defects in III-V Nitrides

Chris G. VAN DE WALLE,[a] Catherine STAMPFL,[a] and Jörg NEUGEBAUER[b]

[a]Xerox Palo Alto Research Center, Palo Alto, California 94304, USA

Tel: +1-650-812-4163; Fax +1-650-812-4140; e-mail: vandewalle@parc.xerox.com

[b]Fritz-Haber-Institut, Abt. Theorie, Faradayweg 4–6, D-14195 Berlin, Germany

Doping problems in GaN and in AlGaN alloys are addressed on the basis of state-of-the-art first-principles calculations. For n-type doping we find that nitrogen vacancies are too high in energy to be incorporated during growth, but silicon and oxygen readily form donors. The properties of oxygen, including DX-center formation, support it as the main cause of unintentional n-type conductivity. For p-type doping we find that the solubility of Mg is the main factor limiting the hole concentration in GaN. We discuss the beneficial effects of hydrogen during acceptor doping. Compensation of acceptors by nitrogen vacancies may occur, becoming increasingly severe as x increases in $Al_xGa_{1-x}N$ alloys.

1. Introduction

Tremendous progress has recently been made in the growth and fabrication of GaN-based electronic and optoelectronic devices [1]. A number of problems still exist, however, which may hamper further progress. Control of doping levels is essential; for n-type doping, the outstanding problems include (i) suppression of background n-type conductivity; (ii) compensation at high doping levels; and (iii) n-type doping of AlGaN alloys. For p-type doping, the main issue concerns increasing the achievable hole concentrations; this requires an understanding of (i) compensation by native defects; (ii) potential metastabilities (as evidenced in persistent photoconductivity); (iii) the role of hydrogen; and (iv) the reasons for increased doping difficulties in AlGaN alloys. In this paper we discuss how a theoretical approach for native defects and dopant impurities, combined with state-of-the-art first-principles calculations, can be used to understand the various factors that govern doping.

After a brief description of the theoretical methods in Section 2, we will discuss our results for n-type doping in Section 3. We will show that nitrogen vacancies play no role in n-type conductivity, and that unintentional n-type doping is usually due to oxygen. We discuss the behavior of oxygen in AlGaN alloys, where a DX transition is predicted to occur. Silicon donors do not exhibit this transition, and it is also absent in the zinc-blende phase. We will also discuss gallium vacancies, which act as compensating centers in n-type GaN, and which are the most likely source of the "yellow luminescence."

Section 4 will address p-type doping, which is now routinely performed in GaN with Mg acceptors and using a post-growth activation step in material grown by MOCVD (metal-organic chemical vapor deposition). Hydrogen plays a crucial role in this process. However, p-type doping levels are still lower than desirable for low-resistance cladding layers and ohmic contacts. We will show that Mg solubility is the determining factor limiting the hole concentration in GaN; incorporation of Mg on interstitial sites or on substitutional nitrogen sites is found to be unfavorable. We will also discuss the prospects of other acceptor impurities for achieving higher doping levels. Some degree of compensation by nitrogen vacancies occurs, and we will discuss the effect of increasing x in $Al_xGa_{1-x}N$ alloys on the doping efficiency. Section 5 concludes the paper.

2. Methods

The equilibrium concentration of impurities or native defects can be expressed as

$$c = N_{\text{sites}} \exp^{-E^f/k_B T} \tag{1}$$

Where E^f is the *formation energy*, N_{sites} is the number of sites the defect or impurity can be incorporated on, k_B is the Boltzmann constant, and T the temperature. Equation (1) shows that defects with a *high* formation energy will occur in *low* concentrations.

The formation energy is not a constant but depends on the growth conditions. For example, the formation energy of an oxygen donor is determined by the relative abundance of O, Ga, and N atoms, as expressed by the chemical potentials μ_O, μ_{Ga} and μ_N, respectively. If the O donor is charged (as is expected when it has donated its electron), the formation energy depends further on the Fermi level (E_F), which acts as a reservoir for electrons. Forming a substitutional O donor requires the removal of one N atom and the addition of one O atom; the formation energy is therefore:

$$E^f(\text{GaN:O}_N^q) = E_{\text{tot}}(\text{GaN:O}_N^q) - \mu_O + \mu_N + qE_F \tag{2}$$

where $E_{\text{tot}}(\text{GaN:O}_N^q)$ is the total energy derived from a calculation for substitutional O, and q is the charge state of the O donor. Similar expressions apply to other impurities and to the various native defects. We refer to Refs. [2] and [3] for a more complete discussion of formation energies and their dependence on chemical potentials.

The quantity E_{tot} in Eq. (2) is obtained from state-of-the-art first-principles calculations that do not require any adjustable parameters or any input from experiment. The computations are founded on density-functional theory, using a supercell geometry and soft *ab initio* pseudopotentials. The effect of d electrons in GaN is taken into account either through the so-called non-linear core correction or by explicit inclusion of the d electrons as valence electrons; the latter proved to be necessary for obtaining accurate results in certain cases [4]. References and further details of the computational approach can be found in Refs. [2], [5], and [6].

The Fermi level E_F is not an independent parameter, but is determined by the condition of charge neutrality. In principle we can write equations such as (2) for every native defect and impurity in the material, and then solve the complete problem (including free-carrier concentrations in valence and conduction bands) self-consistently, imposing charge neutrality. However, it is instructive to plot formation energies as a function of E_F in order to examine the behavior of defects and impurities when the doping level changes. As for the chemical potentials, these are variables which depend on the details of the growth conditions. For clarity of presentation, we set these chemical potentials to fixed values in the figures shown below; however, a general case can always be addressed by referring back to Eq. (2). The fixed values we have chosen correspond to Ga-rich conditions [$\mu_{\text{Ga}} = \mu_{\text{Ga(bulk)}}$], and to maximum incorporation of the various impurities, with solubilities determined by equilibrium with Ga_2O_3, Si_3N_4, and Mg_3N_2.

3. n-type doping

For a long time the nitrogen vacancy was thought to be the source of n-type conductivity in GaN. Our first-principles investigations, however, indicate that nitrogen vacancies are high-energy defects in n-type GaN [2]. Nitrogen vacancies (V_N) do behave as donors; if they are purposely created, for instance during ion implantation, they will increase the electron concentration. However, their high formation energy makes it very unlikely that they would form during growth of undoped or n-type GaN, and hence they cannot be responsible for n-type conductivity. Instead, we have proposed that unintentional impurities such as oxygen and silicon are the actual cause of

the observed unintentional n-type doping [7]. These impurities are calculated to be shallow donors with high solubilities. Figure 1 summarizes our results for native defects and impurities relevant for n-type doping. It is clear that O and Si have much lower formation energies than V_N; these impurities can be readily incorporated in n-type GaN. Both O and Si form shallow donors in GaN. The slope of the lines in Fig. 1 indicates the charge state of the defect or impurity [see Eq. (2)]: Si_{Ga}, O_N, and V_N all appear with slope +1, indicating single donors.
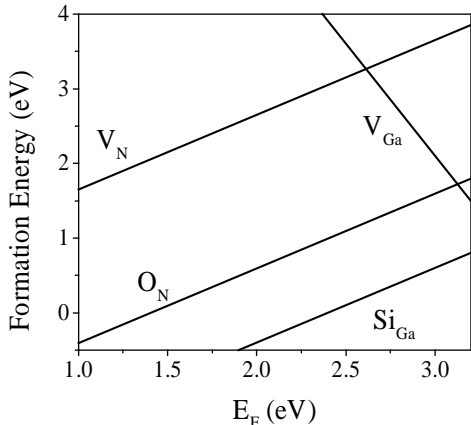

Figure 1: Formation energy vs. Fermi energy for native defects (nitrogen and gallium vacancies) and donors (oxygen and silicon) in GaN. The zero of Fermi energy is located at the top of the valence band.

A few reports had proposed oxygen as a potential source of n-type conductivity in GaN [8, 9]. Still, the prevailing conventional wisdom, attributing the n-type behavior to nitrogen vacancies, proved hard to overcome. Recent experiments have now confirmed that unintentionally doped n-type GaN samples contain Si or O concentrations high enough to explain the electron concentrations [10, 11]. High levels of n-type conductivity have always been found in GaN bulk crystals grown at high temperature and high pressure [12]. Recent high-pressure studies have established that the characteristics of these samples are very similar to epitaxial films which are intentionally doped with oxygen [13, 14]. The n-type conductivity of bulk GaN can therefore be attributed to unintentional oxygen incorporation. The high-pressure experiments have also shown that freeze-out of carriers occurs at pressures exceeding 20 GPa [12, 14, 15]. Originally this observation was interpreted as consistent with the presence of nitrogen vacancies, since the V_N donor gives rise to a resonance in the conduction band, which emerges into the band gap under pressure. However, the observations are also entirely consistent with a "DX-like" behavior of the oxygen donor.

The prototype DX center is Si in GaAs, which undergoes a transition from a shallow to a deep center when hydrostatic pressure is applied, or in AlGaAs alloys [16]. The transition is accompanied by a strong relaxation of the impurity off the substitutional site [17]. We have carried out extensive calculations for oxygen in GaN under pressure, as well as in AlGaN alloys [18]. Under compression the oxygen impurity assumes an off-center configuration: a large outward relaxation introduces a deep level in the band gap. This behavior explains the carrier freezeout in GaN under pressure. Silicon donors do not exhibit this transition, consistent with experiment [15]. Alloying with AlN increases the band gap similar to the application of hydrostatic pressure; one therefore expects that the behavior of the impurities in AlGaN would be similar to that in GaN under pressure. Explicit calculations for oxygen in AlN indeed produce DX behavior [18]. These results are consistent with the observed decrease in n-type conductivity of unintentionally doped $Al_xGa_{1-x}N$ as $x > 0.4$ [19].

Interestingly, we find that the DX transition does *not* occur in *zinc-blende* (ZB) AlGaN. This difference is surprising, since the local environment of the impurity is very similar (the two phases only differ at the positions of third-nearest neighbors and beyond, and no qualitative differences had been observed for any defects or impurities so far [4, 20]. We have explained the difference in DX

behavior by analyzing the interaction between the oxygen impurity and third-nearest neighbors, showing that a Coulombic attraction provides a driving force for the large lattice relaxation [18].

Finally, we note in Fig. 1 that gallium vacancies (V_{Ga}^{3-}) have relatively low formation energies in highly doped n-type material (E_F high in the gap); they could therefore act as compensating centers. Yi and Wessels [21] have found evidence of compensation by a triply charged defect in Se-doped GaN. We have also proposed that gallium vacancies are responsible for the "yellow luminescence" (YL) in GaN, a broad luminescence band centered around 2.2 eV [22]. The origins of the YL have been extensively debated; as discussed in Refs. [22] and [23], the calculated properties of the gallium vacancy are in good agreement with experimental results.

4. p-type doping

p-type doping levels in GaN and AlGaN alloys are still lower than desirable for low-resistance cladding layers and ohmic contacts. Achieving higher hole concentrations with Mg as the dopant has proved difficult. Our investigations [24] have revealed that the determining factor is the solubility of Mg in GaN, which is limited by competition between incorporation of Mg acceptors and formation of Mg_3N_2. It would be interesting to investigate experimentally whether traces of Mg_3N_2 can be found in highly Mg-doped GaN. Mg prefers the substitutional Ga site, and incorporation of Mg on substitutional N sites (Mg_N) or on interstitial sites (Mg_i) was found to be unfavorable. These features are illustrated in Fig. 2.
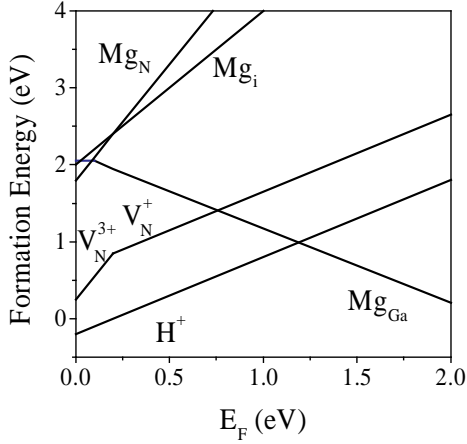

Figure 2: Formation energy as a function of Fermi level for Mg in different configurations (Ga-substitutional, N-substitutional, and interstitial configuration). Also included are the native defects and interstitial H.

Hydrogen has strong effects on the properties of p-type GaN. Many growth techniques, such as MOCVD or hydride vapor phase epitaxy (HVPE) introduce large concentrations of hydrogen in the growing material. The behavior of hydrogen in GaN was analyzed in detail in Refs. [20] and [25]. We found that hydrogen incorporates much more readily in p-type than in n-type GaN. In p-type GaN H behaves as a donor (H$^+$), compensating acceptors. Hydrogen can bind to the Mg acceptors with a binding energy of 0.7 eV. The structure of the resulting complex is unusual in that H does not sit next to the Mg, but binds to a N atom which is a neighbor of the acceptor. As a direct consequence the vibrational frequency of the complex is *not* representative of a Ga-H bond, but rather of a N-H bond. The calculated vibrational frequency (in the harmonic approximation) is 3360 cm^{-1}. Anharmonic effects may lower this frequency by as much as 170 cm^{-1} [26]. Götz *et al.* [27] have reported a value of 3125 cm^{-1} for this local vibrational mode.

The nitrogen vacancy, which had a high formation energy in n-type GaN (see Fig. 1), has a significantly lower formation energy in p-type material, and could act as a compensating center. Figure 2 shows that V_N can occur in a 3+ as well as a + charge state; the +/3+ transition is characterized by a large lattice relaxation [3]. Compensation by V_N may therefore be responsible

for the observed persistent photoconductivity effects [28]. The metastability is associated with the different position of the A_1 state near the valence band in the +1 and +3 charge states; this state is occupied with two electrons the +1 charge state, and empty for the 3+ charge state. The nitrogen vacancy also may give rise to the blue lines (around 2.9 eV) commonly observed by photoluminescence in Mg-doped GaN.

We have carried out similar calculations for the formation energy of native defects and impurities in AlN. The nitrogen vacancy has a strikingly lower formation energy in AlN than in GaN; it will therefore introduce more severe compensation. We propose that compensation by nitrogen vacancies is the likely cause of the decreased doping efficiency of Mg when the Al content is raised in $Al_xGa_{1-x}N$ alloys.

Figure 2 also shows that hydrogen, when present, has a formation energy much lower than that of the nitrogen vacancy. In growth situations where hydrogen is present (such as MOCVD or HVPE) Mg-doped material will preferentially be compensated by hydrogen, and compensation by nitrogen vacancies will be suppressed. The presence of hydrogen is therefore beneficial – at the expense, of course, of obtaining material that is heavily compensated by hydrogen! Fortunately, the hydrogen can be removed from the active region by treatments such as low-energy electron-beam irradiation [29] or thermal annealing [30]. A more complete discussion of the role of hydrogen in GaN is given in Refs. [20] and [25].

For Mg, we thus conclude that achievable doping levels are mainly limited by the solubility of Mg in GaN. We have investigated other candidate acceptors, and evaluated them in terms of solubility, shallow *vs.* deep character, and potential compensation due to incorporation on other sites [31]. None of the candidates exhibited characteristics exceeding those of Mg. In particular, we perceive no noticeable advantage in the use of Be, which has been suggested as a superior dopant.

Finally, we note the importance of avoiding oxygen contamination during growth of p-type GaN. The oxygen formation energy shown in Fig. 1 clearly extrapolates to very low values in p-type GaN. Any oxygen present in the growth system will therefore be readily incorporated during p-type growth. In addition, complex formation between oxygen and magnesium can make oxygen incorporation even more favorable: we find that the Mg-O complex has a binding energy of 0.6 eV in both GaN and AlN.

5. Conclusions

We have presented a comprehensive overview of our current understanding of point defects and dopant impurities in nitride semiconductors, based on first-principles calculations. Our main conclusions for n-type GaN are that (i) nitrogen vacancies are *not* responsible for unintentional n-type conductivity; (ii) Si and O donors can be incorporated in large concentrations, likely causing unintentional n-type doping; (iii) oxygen (but not silicon) behaves as a DX center in GaN under pressure and in AlGaN alloys; (iv) the DX transition does not occur in zinc-blende material; and (v) gallium vacancies are the likely source of the yellow luminescence. For p-type GaN we found that (i) the hole concentration obtained with Mg doping is limited due to Mg solubility; (ii) incorporation of Mg on interstitial sites or antisites is not a problem; (iii) hydrogen has a beneficial effect on p-type doping because it suppresses compensation and enhances acceptor incorporation; (iv) compensation by nitrogen vacancies may occur, becoming worse as x increases in $Al_xGa_{1-x}N$; and (v) other acceptor impurities do not exhibit characteristics superior to Mg.

References

[1] For a recent review, see F. A. Ponce and D. P. Bour, Nature **386** (1997) 351.

[2] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50** (1994) 8067.

[3] J. Neugebauer and C. G. Van de Walle, in Festkörperprobleme/Advances in Solid State Physics, Vol. **35**, ed. R. Helbig (Vieweg, Braunschweig/Wiesbaden, 1996) p. 25.

[4] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **339** (1994) 687.

[5] R. Stumpf and M. Scheffler, Comp. Phys. Commun. **79** (1994) 447.

[6] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **408** (1996) p. 43.

[7] J. Neugebauer and C. G. Van de Walle, Proc. ICPS-22 (World Scientific, Singapore, 1995) p. 2327.

[8] W. Seifert, R. Franzheld, E. Butter, H. Sobotta, and V. Riede, Cryst. Res. & Technol. **18** (1983) 383.

[9] B.-C. Chung and M. Gershenzon, J. Appl. Phys. **72** (1992) 651.

[10] W. Götz, N. M. Johnson, C. Chen, H. Liu, C. Kuo, and W. Imler, Appl. Phys. Lett. **68** (1996) 3114.

[11] W. Götz , J. Walker, L. T. Romano, and N. M. Johnson, Proc. Mater. Res. Soc. Symp. **449** (1997) 525.

[12] P. Perlin, T. Suski, H. Teisseyre, M. Leszczyński, I. Grzegory, J. Jun, S. Porowski, P. Boguslawski, J. Bernholc, J. C. Chervin, A. Polian, and T. D. Moustakas, Phys. Rev. Lett. **75** (1995) 296.

[13] C. Wetzel, T. Suski, J. W. Ager III, W. Walukiewicz, S. Fisher, B. K. Meyer, I. Grzegory, and S. Porowski, Proc. ICPS-23 (World Scientific, Singapore, 1996), p. 2929.

[14] P. Perlin, T. Suski, A. Polian, J. C. Chervin, W. Knap, J. Camassel, I. Grzegory, S. Porowski and J. W. Erickson, Proc. Mater. Res. Soc. Symp. **449** (1997) 519.

[15] C. Wetzel, T. Suski, J. W. Ager III, E. R. Weber, E. E. Haller, S. Fischer B. K. Meyer, R. J. Molnar, and P. Perlin, Phys. Rev. Lett. **78** (1997) 3923.

[16] P. Mooney, in *Deep Centers in Semiconductors*, ed. S. T. Pantelides (Gordon and Breach, Yverdon, Switzerland, 1992) p. 643.

[17] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61** (1988) 873.

[18] C. G. Van de Walle (to be published).

[19] H. G. Lee, M. Gershenzon, and B. L. Goldenberg, J. Electron. Mat. **20** (1991) 621.

[20] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75** (1995) 4452.

[21] G.-C. Yi and B. W. Wessels, Appl. Phys. Lett. **69** (1996) 3028.

[22] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **69** (1996) 503.

[23] C. G. Van de Walle and J. Neugebauer, Proc. Mater. Res. Soc. Symp. **449** (1997) 861.

[24] J. Neugebauer and C. G. Van de Walle, Proc. Mater. Res. Soc. Symp. **395** (1996) 645.

[25] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **68** (1996) 1829.

[26] B. G. Johnson, P. M. W. Gill, and J. A Pople, J. Chem. Phys. **98** (1993) 5612.

[27] W. Götz, N. M. Johnson, J. Walker, D. P. Bour, M. D. McCluskey, and E. E. Haller, Appl. Phys. Lett. **68** (1996) 3725.

[28] C. Johnson, J. Y. Lin, H. X. Jiang, M. A. Khan, and C. J. Sun, Appl. Phys. Lett. **68** (1996) 1808; J. Z. Li, J. Y. Lin, H. X. Jiang, A. Salvador, A. Botchkarev, and H. Morkoc, *ibid.* **69** (1996) 1474.

[29] H. Amano, M. Kito, K. Hiramatsu, and I. Akasaki, Jpn. J. Appl. Phys. **28** (1989) L2112.

[30] S. Nakamura, N. Iwasa, M. Senoh, and T. Mukai, Jpn. J. Appl. Phys. **31** (1992) 1258.

[31] J. Neugebauer and C. G. Van de Walle, Proc. ICPS-23 (World Scientific, Singapore, 1996) p. 2849.

Invited talk at the Second International Conference on Nitride Semiconductors, Tokushima, Japan, October 27-31, 1997; accepted for publication in J. Cryst. Growth.